Finding Equilibrium in Multi-Agent Games with Payoff Uncertainty

Wenshuo Guo  Mihaela Curmei  Serena Wang  Benjamin Recht  Michael I. Jordan
Two-player zero-sum game: Rock-Paper-Scissors

- Player 1 and Player 2
- Zero sum:
  - **WIN:** Player 1 gets 1 point, Player 2 loses 1 point
  - **LOSE:** Player 1 loses 1 point, Player 2 gets 1 point
- Pure strategy: e.g. always plays Rock
- Mixed strategy: $x = (x_1, x_2, x_3) \in \Delta$
  $$y = (y_1, y_2, y_3) \in \Delta$$

**Nash Equilibrium:** A pair of strategies $(x^*, y^*)$, s.t.
$$x^* \in \arg \max_{x \in \Delta} x^T Ay^* \quad y^* \in \arg \max_{y \in \Delta} (x^*)^T (-A)y$$
Games with incomplete information

● With **complete game information**:
  ○ Nash equilibria for **general-sum** finite games can be intractable to compute
  ○ Nash equilibrium of a two-player **zero-sum** game can be computed efficiently

● In practice, the parameters of a game are often only **partially known**
  ○ Companies get uncertain profit
  ○ Bidders have uncertain gain ... etc

How can we find equilibria in games in a way that is **robust to uncertainty** in the game parameters?

This work: payoff uncertainty
Two-player zero-sum game: Rock-Paper-Scissors

- Player 1 and Player 2
- Zero-sum
- Uncertain payoff matrix: $A \in \mathcal{U}$

If Player 1 wins: $\text{win} \in [0, 1]$ ~ uncertainty set

If Player 1 loses: $\text{loss} \in [-1, 0]$ ~ uncertainty set

Goal: Nash Equilibria for both players remains to be a Nash for any feasible payoff. If such strategies exist, they are known as *ex-post equilibrium*. 

W. Guo, M. Curmei, S. Wang, B. Recht, M. Jordan. University of California, Berkeley
Ex-post equilibrium

Suppose $A \in \mathcal{U}$ where $\mathcal{U}$ is some uncertainty set.

A pair of strategies $x^*, y^*$ is at an ex-post equilibrium if:

$$
\begin{align*}
    x^* &\in \text{arg max}_{x \in \Delta_n} x^T Ay^* \\
    y^* &\in \text{arg max}_{y \in \Delta_m} x^* T (\mathbf{-} A)y
\end{align*}
\quad \forall A \in \mathcal{U}
$$

In this work: (“distribution-free” uncertainty) We consider uncertainty sets that correspond to convex combinations of a set of known matrices:

$$
A \in \mathcal{U} = \text{conv}(\{A_1, \ldots, A_k\})
$$
Contribution of this work

- In zero-sum one-shot games:
  - Existing ex-post equilibrium of zero-sum polymatrix games can be found efficiently using linear programming

- In repeated stochastic games:
  - Ex-post equilibrium $\rightarrow$ Ex-post Markov perfect equilibrium (MPE)
  - Sufficient conditions for the existence of such an equilibrium
  - Any feasible values of a two-player zero-sum stochastic game can be bounded in a tight interval using dynamic programming
One-shot game: Existence of Ex-post equilibrium

- Ex-post equilibrium doesn’t always exist
  - It exists only if the set of Nash Equilibria for all possible payoff matrices contains a common strategy.
- Example:
  - Having uncertainty in the upper right corner would mean that Player 1 would have a bias against playing Rock, thus the equiprobable strategy is no longer at Equilibrium.

Result: Any existing ex-post equilibrium can be found via Linear Programming for certain zero-sum games with bounded polyhedral uncertainty sets.
Example: ex-post equilibrium exists

As a concrete example, if for all payoff matrices in the uncertainty set, the payoffs for winning and losing in are all equal in magnitude, then $x^* = y^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ is an ex-post equilibrium.

$$U = \text{conv}(\{ \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \})$$
Example: ex-post equilibrium does not exist

If we are only uncertain about how much Player 1 would win in a Rock-Scissor setting, having $W<1$ would mean that Player 2 would have a bias for playing Rock, thus the equiprobable strategy is no longer at Equilibrium.

$$\mathcal{U} = \text{conv}({ 0 -1 1, 1 0 -1, -1 1 0, 0 -1 0, 1 0 -1, -1 1 0 })$$

**Further:** We provide characterizations of the maximal uncertainty sets for an ex-post equilibrium to exist. *(details in the paper)*
Solving for existing ex-post equilibrium

**LP(1)**

\[
\begin{align*}
\min_{x,y} & \quad \sum_{i=1}^{k} w\{x,i\} + w\{y,i\} \\
\text{s.t.} & \quad w\{x,i\} \geq -e_j^T A_i y \quad \forall j = 1, \ldots, n; i = 1, \ldots, k \\
& \quad w\{y,i\} \geq x^T A_i e_l \quad \forall l = 1, \ldots, m; i = 1, \ldots, k \\
& \quad x^T 1 = 1, x \geq 0 \\
& \quad y^T 1 = 1, y \geq 0
\end{align*}
\]

**Theorem I (simplified two-player case)**

Any ex-post equilibrium \((x^*, y^*)\) for a two-player zero sum with uncertain payoff matrix \(A \in \mathcal{U} = \text{conv}(\{A_1, \ldots, A_k\})\) gives an optimal solution to LP(1).

Conversely, any optimal solution to LP(1) is an ex-post equilibrium of the game.
Extending to multi-players

Theorem 1 holds for more general N-player polymatrix games:

- Each pair of players play a two-player matrix game
- Each player chooses a single strategy, which is applied to all his pairwise games
- Zero-sum: The sum of the payoffs for all players $= 0$
Extending to stochastic games with infinite time horizon

**Stochastic game**: Game is played over and over. Each stage has a **stage payoff matrices**, with transition probability between stages given by $P(\text{new stage} | \text{current stage, players’ actions at current stage})$.

**Example**: the streaming services’ actions in **Stage 1** change user behavior in **Stage 2**.
Ex-post Markov perfect equilibrium (simplified two-player case)

Player 1’s expected discounted payoff: (2 players with strategies $x, y$)

$$\Pi_1(x, y; s^0) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t M_1(x(s^t, \mathcal{H}^t), y(s^t, \mathcal{H}^t); s^t)\right]$$

**Ex-post MPE:** A set of strategies $(x^*(s), y^*(s))$ is an ex-post MPE of the stochastic game iff it maximizes the expected discounted payoff for all the players under all feasible stage payoff matrices $A(s) \in \mathcal{U}(s), \forall s \in S$.

**Results:** (applies to N-player)
- **Sufficient conditions** for the existence of such an ex-post MPE;
- **Value of the game** is achieved at MPE. It can be **bounded up to a tight interval** using dynamic programming.
Future directions

- **Necessary conditions** for the existence of ex-post equilibria, and ex-post MPE
- Games with **more general payoff structures** (beyond polymatrix) and **uncertainty sets** (beyond polyhedral)
- Uncertain payoffs: **potential relations to practical reinforcement learning problems with proxy rewards**
- **Other sources of uncertainties** beyond payoff: transition probabilities

Thank you for your time!

For more details, please see the paper at: https://arxiv.org/pdf/2007.05647.pdf